CODIP

Technology at Work

AT&T Government Solutions, Inc.

Patrick Emery

Lewis Hart

PatEmery@att.com or LewisHart@att.com

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
AUG 2004		**00-00-2004 to 00-00-2004**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
CODIP Technology at Work	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **AT&T Government Solutions Inc,1900 Gallows Rd,Vienna,VA,22182**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Proceedings of the 2004 Performance Metrics for Intelligent Systems Workshop (PerMIS ?04), Gaithersburg, MD on August 24-26 2004

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**10**	

Overview
Key Ideas

The CODIP program provides frameworks and components for intelligent processing of information based on its semantics.

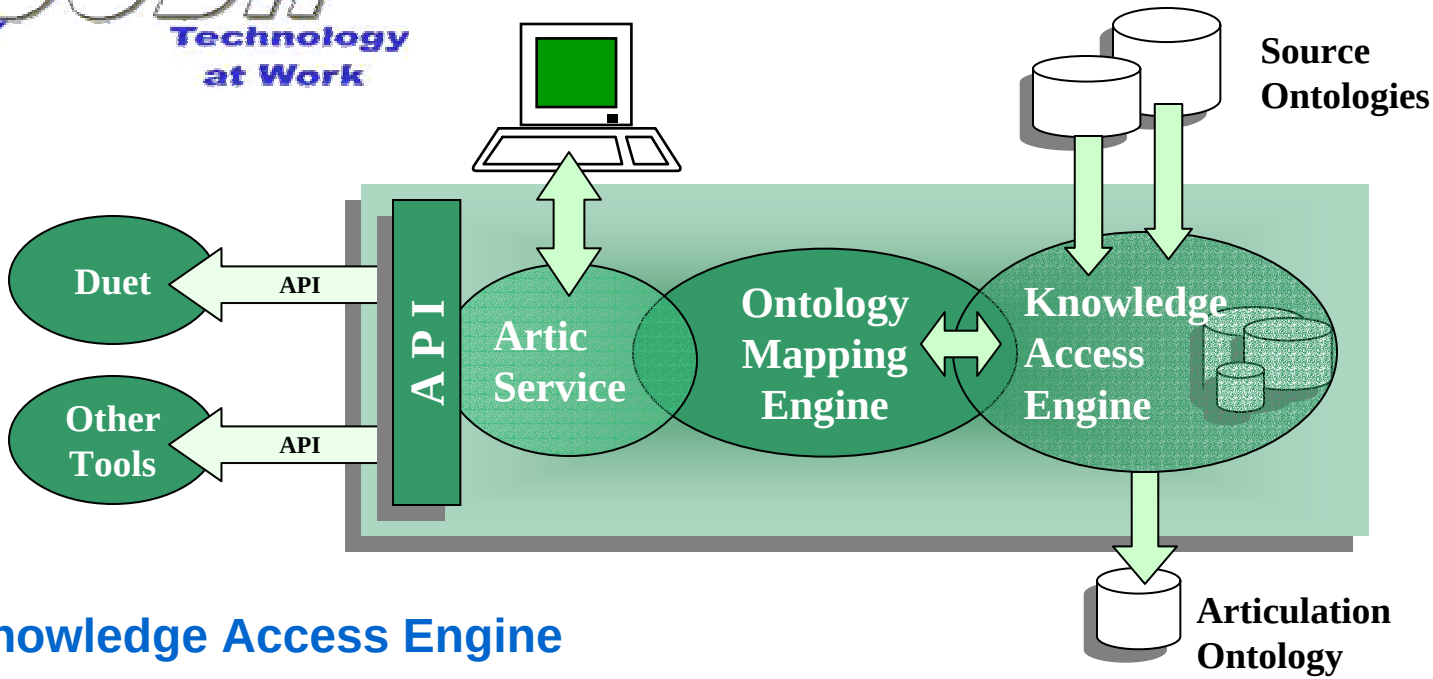
- **Distribution of information from publishers to subscribers using subscriber defined semantic queries.**
- **Automatic generation of semantic mapping between ontologies to facilitate database integration, content translation and distribution.**
- **Application of a UML technology to leverage existing resources to provide knowledge engineering capability.**
- **Ontological processing components and services that can bring built-in knowledge processing capability to applications.**

Overview
Applications and Products

- **Primary products support these applications:**
  - **Duet** to support visualization, application and management of ontologies using the UML/MOF engineering standards,
  - **Kage** to support applications with analysis, translation, and repository functionality,
  - **ODKD** for semantics based publication of information to subscribers, and
  - **Artic** to support using multiple ontologies concurrently by finding and codifying relationships between their concepts.
- **These products are built from library of reusable components that may be integrated into other applications.**

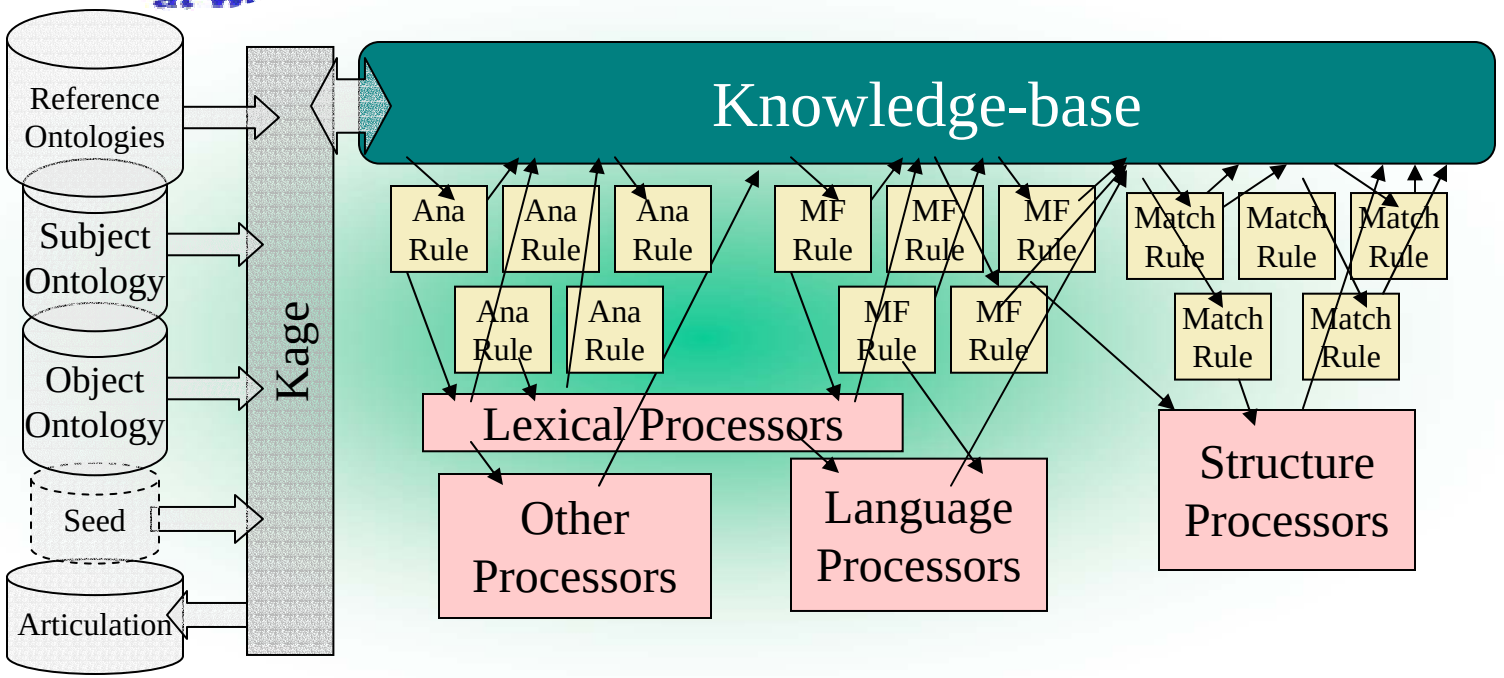
http://codip.grci.com

Artic Functional Architecture



- **Knowledge Access Engine**
  - Provides management and access to collections of ontologies.
- **Ontology Mapping Engine**
  - Provides automated analysis of potential mappings between ontologies and builds articulation ontologies that codify the mappings.
- **Artic Service**
  - Provides APIs, command line and web based access to the mapping engine.

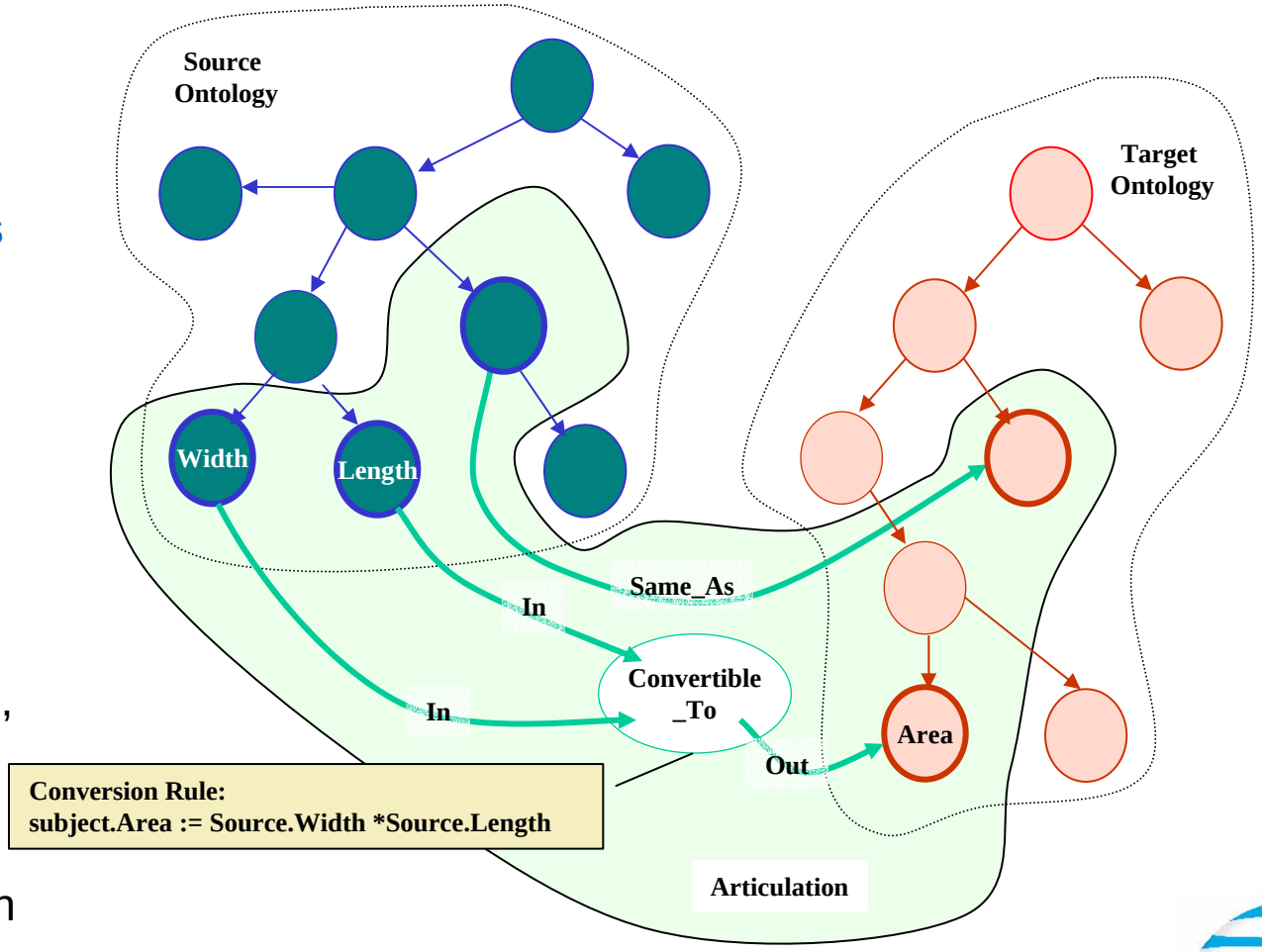
Artic Ontology Mapping Engine



- **Multiple Layers**
  - Non-procedural rules and procedural processors, invoked by rules.
- **Multiple Phases**
  - Analysis, Match Factors, Matches.

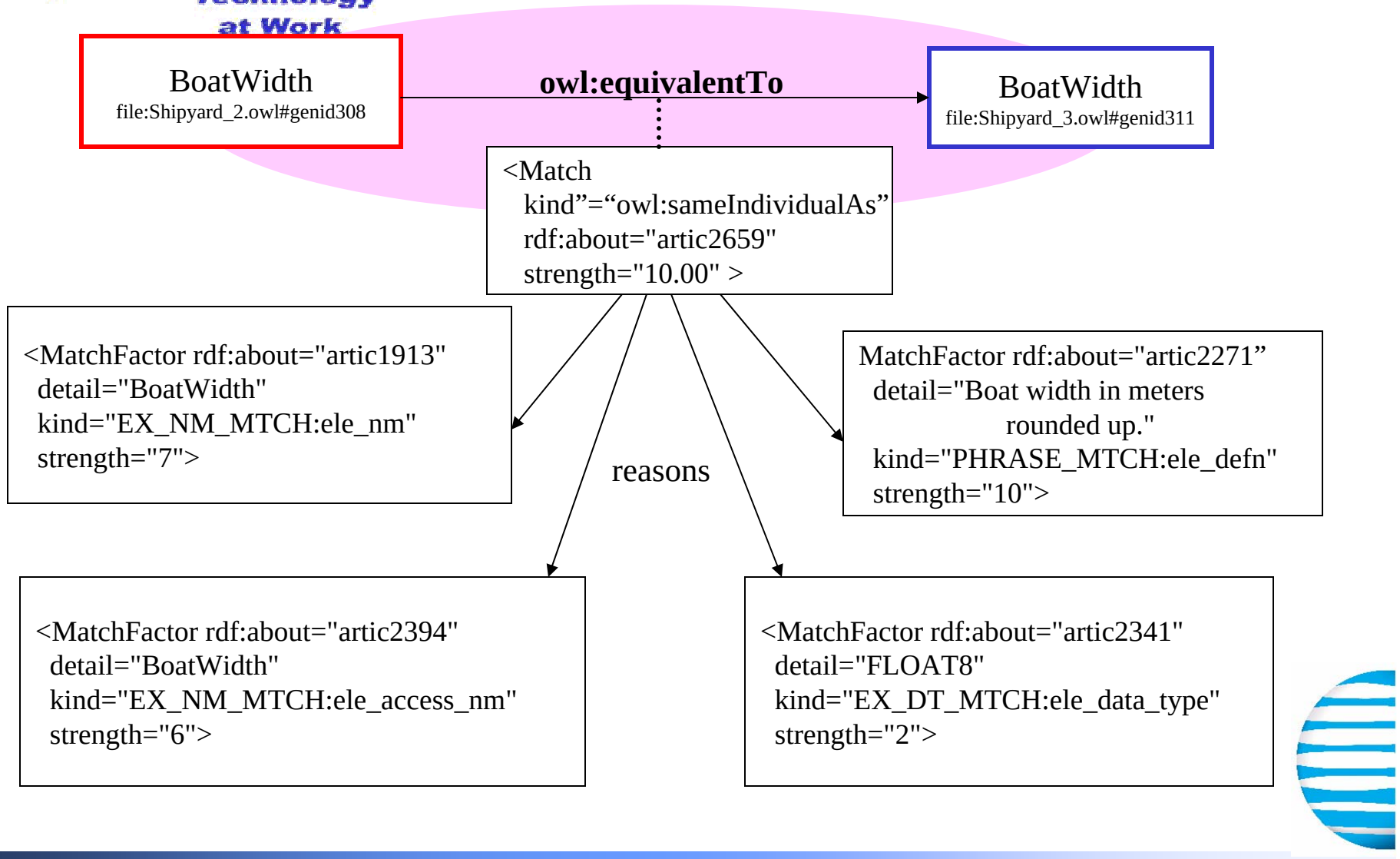
Articulation Ontologies

- **Articulations are specialized ontologies that relate concepts in other ontologies.**
  - Relationships of various types:
    - » Similarity, Part-Of, Kind-Of, Temporal, Spatial, and Domain Specific.
  - Multiplicity may be 1:1, 1:M, M:1 or M:M
  - Variable 'strength'.
  - May include conversion rules, which may be one-way.



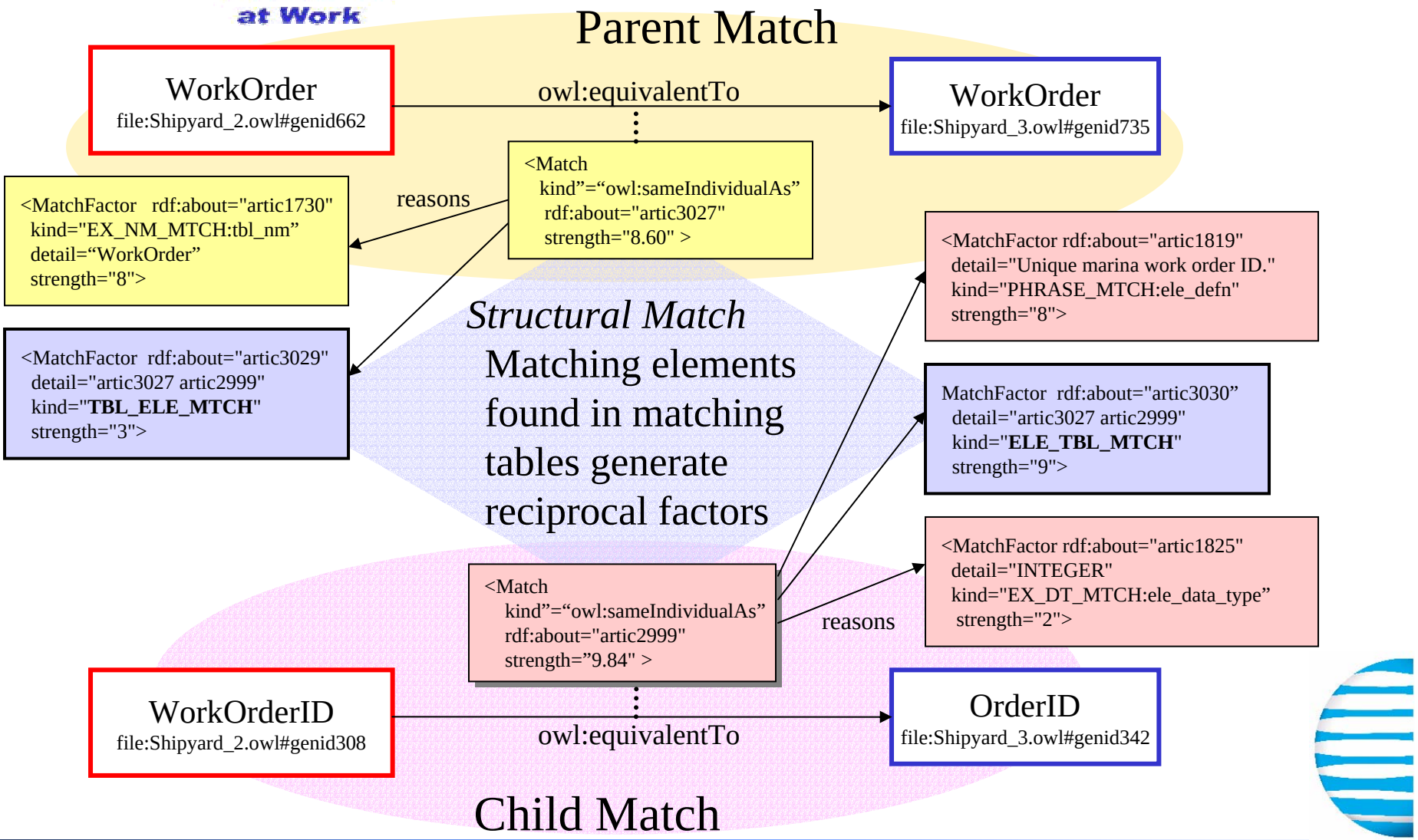
Example Match

Simple, 'Perfect'

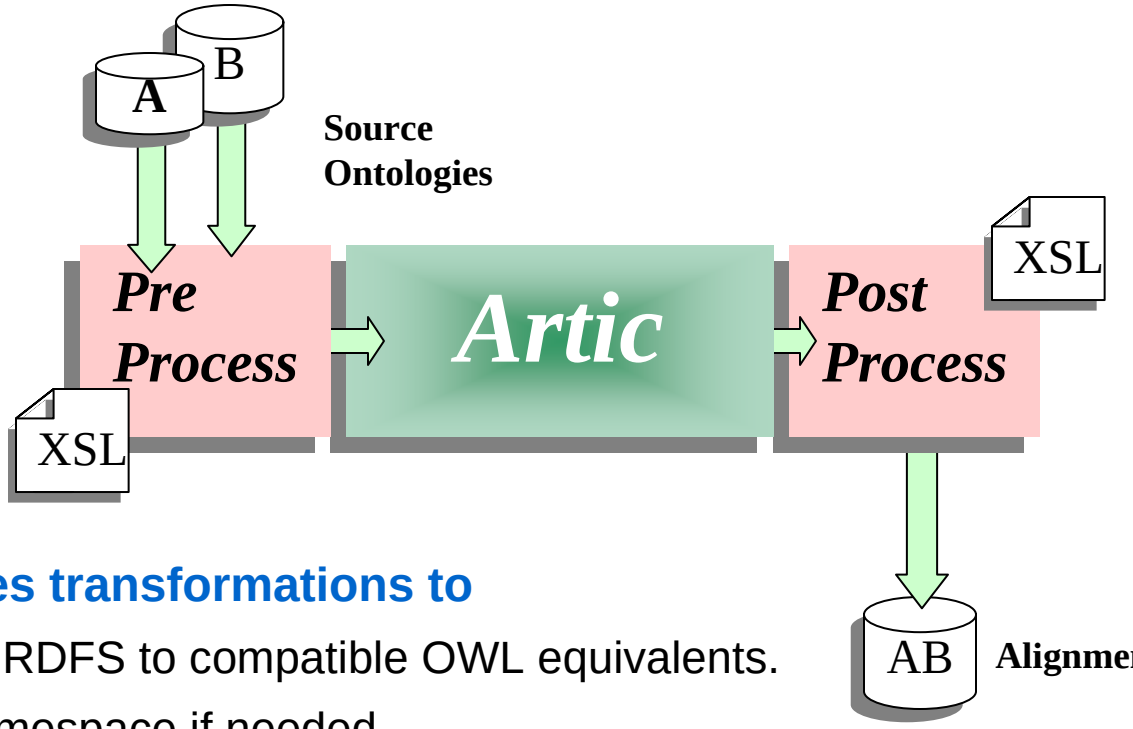


Articulation Example

Structural, Imperfect Match



I3Con Processing



- **Pre Processing applies transformations to**
  - Convert DAML and RDFS to compatible OWL equivalents.
  - Adds XML Base namespace if needed.
- **Post Processing applies transformations to**
  - Convert Articulation to Alignment format.
  - Remove low confidence (< 0.8) and uninteresting matches (e.g. ‘genid’).
  - Removes faulty matches (e.g. rdf:ID=“ “)

I3Con Results Summary

	Animals	Sports	Comsci	Hotel	Network	P&P	P&P-noi	Russia
Alignments	83	187	0	15	44	128	286	150
Confidence								
Maximum	1.0006	0.9671	0.0000	0.9396	1.0011	1.0012	0.9824	0.9997
Minimum*	0.8423	0.8423	0.0000	0.8435	0.8423	0.8423	0.8423	0.8423
Average	0.9283	0.9223	N/A	0.8808	0.9051	0.9111	0.9255	0.9024
Median	0.9326	0.9384	N/A	0.8709	0.9374	0.9374	0.9358	0.9359

* Confidence < 0.8 removed.

- **Issues**
  - Namespaces – XML Base needed to allow local file usage.
  - ID verses rdf:ID – leads to resources with no ID.
- **Semantic differences between RDF/S, DAML+OIL, OWL**
- **Some results not understood**
  - Comsci topic lead to no alignments
  - Removal of instance data in People&Pets produced more alignments.